韓國컴퓨터情報學會 論文誌
第14卷 第8號, 2009. 8.

2009-14-8-2-4

시공간 정보를 이용한 자막 탐지 및 향상 기법

정 종 면*

A Method for Text Detection and Enhancement using Spatio-Temporal Information

Jongmyeon Jeong*

요 약

디지털 비디오에서 텍스트 정보는 비디오 데이터의 시청각적인 정보를 보강하고 부가 정보를 제공하기 때문에 방대한 멀티미디어의 내용을 예측할 수 있는 중요한 단서를 제공한다. 본 논문에서 제안된 방법은 주어진 영상열로부터 자막의 획 특징을 이용하여 자막을 탐지하고, 프로젝션을 이용하여 자막의 위치를 찾는다. 찾아진 자막을 포함하는 바운딩박스에 대한 기하학적인 검증을 거친 후, 서로 인접하는 프레임에 있는 바운딩박스 중 공간적으로 동일한 위치의 바운딩박스에 대한 MAD를 이용하여 바운딩박스를 추적하고, 시간적 중복성을 이용하여 바운딩박스 영역의 화질을 향상시킨다. 다양한 비디오에 대한 실험 결과는 제안된 방법의 타당성을 보인다.

Abstract

Text information in a digital video provides crucial information to acquire semantic information of the video. In the proposed method, text candidate regions are extracted from input sequence by using characteristics of stroke and text candidate regions are localized by using projection to produce text bounding boxes. Bounding boxes containing text regions are verified geometrically and each bounding box existing same location is tracked by calculating matching measure, which is defined as the mean of absolute difference between bounding boxes in the current frame and previous frames. Finally, text regions are enhanced using temporal redundancy of bounding boxes to produce final results. Experimental results for various videos show the validity of the proposed method.

▸ Keyword : 자막 탐지(text detection), 자막 찾기(text localization), 자막 향상(text enhancement), 모폴로지(morphology)

• 제1저자 : 정종면
• 투고일 : 2009. 07. 06, 심사일 : 2009. 07. 17, 게재확정일 : 2009. 08. 19.
* 국립 목포해양대학교 해양전자통신공학부 교수
※ 이 논문은 2006년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(KRF-2006-331- D00517)

Ⅰ. 서 론

1.1 배경

컴퓨터 기술과 멀티미디어 처리 기술의 급속한 성장은 각종 정보를 시청각화하여 제공함으로써 대량의 데이터에 대한 효율적인 사용을 가능하게 한다. 나날이 증가하고 있는 멀티미디어 컨텐트에 대한 수요는 멀티미디어 컨텐트의 폭발적인 증가로 이어졌고, 이에 따라 이 방대한 정보에 대한 효율적인 관리에 대한 다양한 요구가 나타나게 되었다[1].

멀티미디어 컨텐트에 흔히 나타나는 자막 정보는 비디오 데이터들의 시청각적인 정보를 보강하고 부가 정보들을 사용자에게 제공하며 방대한 멀티미디어 데이터의 내용을 예측할 수 있는 중요한 단서를 제공할 수 있다. 또한 방대한 양의 비디오 데이터로부터 사용자가 원하는 데이터를 효율적으로 제공하기 위한 내용기반 비디오 검색 시스템을 구축할 때 자막 정보를 이용하는 것은 비디오의 컬러, 모양, 질감, 움직임 등의 특징을 추출하고 인식하는 것에 비하여 비디오로부터 특징 추출하여 인식하기 위한 비용과 정확성이 더 우수하고 보다 사용자가 원하는 데이터를 찾는데 용이하다. 따라서 주어진 비디오로부터 자막 정보를 추출하여 인식할 수 있다면 비디오 색인 및 검색에 효과적으로 사용될 수 있다.

1.2 기존의 연구

주어진 비디오로부터 텍스트를 추출하기 위한 수 많은 연구가 오래 전부터 이루어져 왔으며 많은 성과가 있었다[2-7]. 기존의 연구에서 정립된 자막 추출 알고리즘들은 동영상에 자막이 존재하는지 여부를 판단하는 자막 프레임 추출(text detection) 단계, 자막 프레임이 존재하는 프레임들로부터 자막의 최소 영역(minimum text region) 위치를 찾아내는 단계(text localization), 그리고 추출된 자막 최소 영역에서 자막을 분리(text segmentation)하는 단계로 이루어진다[2].

자막 탐지 단계에서는 자막을 일종의 질감(texture)으로 간주한 다음, 웨이브릿 변환(wavelet transform[8], 이산 여현 변환(discrete cosine transform)[9], 가보필터(Gabor filter)[10], 공간적 변위(spatial variance)[11] 등의 특징을 추출하고, 그 특징을 신경망(neural network), SVM(support vector machine)과 같은 패턴 분류기를 이용하여 자막의 질감과 유사한 질감을 갖는 영역을 탐지하거나, 자막을 일정한 색상을 갖는다고 가정하고 일정한 색상을 갖는 영역을 탐지한 후, 영역의 연결성을 검사하는 방법이 있다[6][12]. 자막 위치 찾기 단계는 주어진 작은 자막 영역들을 통합하여 하나로 병합해 나가는 상향식 접근 방법[12]과, 전체 영역을 분할하여 작은 영역으로 나누어서 자막의 위치를 찾는 하향식 접근 방법이 있다[13-14]. 자막 분리 단계에서는 자막 영역을 이루는 픽셀은 다른 영역과는 다른 색상을 일관되게 갖는다는 가정 하에, 배경과 자막 영역을 분리해 낸다[15]. 한편, 획 정보를 이용하여 씨앗점을 추출한 다음 모폴로지를 적용하여 효과적으로 자막을 추출할 수 있는 알고리즘이 제안되었다[7]. 이 방법은 획이 갖는 구조적 특징을 이용하여 씨앗점을 탐색하고, 이를 모폴로지 연산에 적용함으로써 자막을 추출하였는데, 복잡한 배경을 갖는 비디오에서도 자막을 비교적 잘 추출할 수 있으나, 인공 구조물을 갖는 복잡한 배경에서 오동작하는 (false positive) 문제를 가지고 있다.

한편 비디오의 시간적 정보와 그에 따른 움직임 정보, 그리고 각 프레임간의 중복성 (redundancy)을 이용하여 정지자막 또는 이동자막을 추출하기 위한 노력이 있어왔다[16-17]. Huang 등은 시간적 정보를 이용하여 비디오로부터 이동 텍스트를 탐지하기 위한 알고리즘을 제안하였는데, 주어진 인접한 두 프레임에 대해 주어진 영상을 부블록(sub-block)으로 분할한 다음 블록 매칭을 통해 움직임 벡터를 계산하고, 움직임 벡터가 존재하는 블록 중 에지의 밀도와 에지의 수가 고려중인 스캔라인에 임계치 이상 존재하는 경우 이를 텍스트 후보 영역으로 판정하였다[16]. Palma 등은 비디오로부터 기울어진 텍스트를 포함하는 그래픽 텍스트(graphic text)를 추출하기 위한 알고리즘을 제안하였다[17]. 이 방법은 텍스트 탐지를 위해 Cortez 등이 제안한 분할과 합병(split and merge)알고리즘을 사용하여 입력 영상을 분할 한 후, 문자 정보의 기하학적인 특징을 이용하여 문자(character) 영역만 남기고 나머지 영역을 제거한 다음, 이들 문자들을 인접성, 정렬 상태, 크기, 밝기 등을 고려하여 단어(word)를 생성한다. 생성된 단어들은 움직임 분석을 위해 인접한 다음 프레임에서 생성된 단어들과의 비교를 통해 추적된다.

지난 10여년 동안 텍스트 추출을 위한 많은 연구 성과가 있었으나 다양한 장르의 비디오에 적용하기에는 여전히 많은 어려움이 존재한다. 기존의 연구들은 공통적으로 텍스트 영역은 배경과 높은 대비(contrast)를 갖는다는 점을 가정하고 있으며, 텍스트가 수평 혹은 수직 방향으로 정렬되어 있다는 가정과 텍스트의 크기는 일정하다는 가정 하에 텍스트를 추출한다. 그러나 비디오에 존재하는 텍스트 정보는 그 해상도가 좋지 않고, 복잡한 배경, 다양한 폰트와 특수 효과의 영향 등으로 인해 다양한 장르의 비디오에 적용하기에 어렵다. 즉,

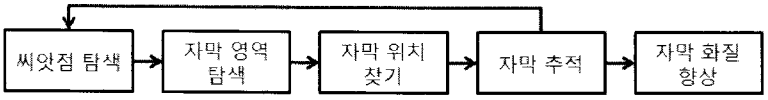


그림 1 시공간 정보를 이용한 자막 탐지 및 자막 향상 알고리즘 블록도
Fig. 1. Block diagram of text detection and enhancement algorithm using spatio-temporal information

질감에 기반한 방법은 자막과는 관계 없는 복잡한 배경이 존재하는 경우 오동작하는 경우가 대부분이며, 방대한 폰트와 다양한 배경, 그리고 자막에 대한 특수 효과 등으로 인해 신경망이나 SVM을 적용하기 위한 training이 쉽지 않다. 또한 색상이나 밝기 정보를 이용한 방법은 대부분의 범용 비디오에서 자막의 색상이 다양하게 나타나기 때문에 자막 영역에 대한 색상정보를 추출하는 것이 선행되어야 한다. 또한 시간 정보를 이용한 Huang 등의 방법[16]은 수행 시간에 대한 고려를 포함하여 빠른 수행시간으로 이동 자막을 탐지할 수 있으나, 카메라의 이동에 의한 배경 변화가 있을 경우 오동작할 가능성이 매우 높다.

1.3 목적 및 논문의 구성

본 논문에서는 참고문헌 [7]에서 제안한 방법으로 자막 영역을 탐지한 다음 비디오의 시공간적인(spatio-temporal) 특징을 이용하여 복잡한 배경을 갖는 비디오로부터 자막의 색상, 위치 등에 무관하게 자막을 강건하게 추출하기 위한 알고리즘을 제안한다. 그림 1은 제안하는 방법의 개략적인 흐름을 보이고 있는데, 그림에서 보이는 바와 같이 주어진 한 개의 프레임으로부터 자막의 존재 여부를 탐지한 후 자막의 위치를 찾아 바운딩박스로 표현한다. 시간 정보를 이용하여 자막 바운딩박스를 추적한 다음, 프레임 사이의 중복성을 이용하여 화질을 향상시켜 최종 결과를 도출한다. 본 논문의 이후 구성은 다음과 같다. II장에서는 참고문헌 [7]에서 제안한 자막 영역 탐지 기법에 대해 설명하고, III장에서는 자막 추적 기법에 대해 설명한다. IV장에서 실험 결과를 보인 후 V장에서 결론을 맺는다.

II. 자막 영역 탐지

본 논문에서 사용하는 자막 영역 탐지 알고리즘은 참고문헌 [7]에서 사용된 방법을 기반으로 한다. 간략하게 이를 기술하면 다음과 같다.

2.1 씨앗점 탐색 및 자막 영역 탐지

자막은 한 개 이상의 획으로 구성되어 있는데, 디지털 이미지에서 획을 이루는 경계선은 밝기값이 크게 변한다. 또한 획은 획의 경계선을 이루는 부분이 서로 쌍을 이루고, 이 쌍은 항상 반대 방향 성분을 갖는다. 즉, 획을 이루는 임의의 에지는 일정한 거리 이내에 자신의 방향과 반대 방향의 에지를 대응쌍으로 갖고, 서로 대응되는 에지 사이의 거리는 획의 폭 크기와 같다. 자막 영역이 아닌 영역에서 나타나는 에지에서 이런 대응관계가 나타날 가능성은 희박하며, 자막 영역에서는 이런 대응관계가 집중적으로 나타난다. 이런 대응관계가 나타나는 픽셀을 후보 씨앗점이라고 했을 때 자막 영역에서는 후보 씨앗점이 집중적으로 나타나지만 그렇지 않은 영역에서는 후보 씨앗점이 집중되지 않는다. 따라서 특정 영역에서 후보 씨앗점의 수가 임계치 이상이면 그 영역의 후보 씨앗점들을 씨앗점으로 하고 그렇지 않은 영역의 후보 씨앗점은 모두 버린다.

씨앗점들이 자막으로부터 얻어진 것이라면 자막 크기 정도의 영역에 집중적으로 나타나며, 자막과 유사한 모양의 영상에서 얻어진 것이 아니라면 고립되어 나타난다. 이런 특징을 이용하여 씨앗점에 대한 모폴로지 연산을 통해 자막후보영역을 탐색한다. 먼저 주어진 씨앗점에 대해 모폴로지 폐쇄 연산(morphological closing)을 수행하면 고립되어 있는 씨앗점, 즉 잘못된 씨앗점들은 변화가 없는 반면, 실제 자막을 이루는 씨앗점들은 일정한 영역에 집중적으로 분포하기 때문에 서로 연결되어 한 개의 영역으로 통합되어 자막 후보 영역이 된다. 그런 다음, 모폴로지 개방 연산(morphological opening) 연산을 수행하면 고립되어 있는 영역은 제거되고, 자막 영역에 포함되어 있던 획 영역은 변화 없이 그대로 존재한다. 따라서 씨앗점에 대해 모폴로지 폐쇄 연산과 개방 연산을 수행한 후에도 남아 있는 영역이 존재한다면 그 영역은 자막영역이 되며, 자막이 탐지되었다고 할 수 있다.

2.2 자막 위치 찾기

복잡한 배경의 영향을 최소화하기 위하여 자막 후보 영역에 존재하는 에지에 대해 4-CC 레이블링을 수행하여 자막 후보 영역에 존재하는 에지 영역들을 서로 분리한 다음, 서로 분리된 에지 영역들 중 씨앗점을 포함하는 영역을 제외하고 모두 제거한다. 이상적인 경우, 이 과정을 거친 에지 영역은

획의 경계선을 이루고 있다고 할 수 있으며, 경계선 영역의 내부는 획 영역이기 때문에 일정한 밝기 분포를 갖는 영역이 존재한다. 자막 영역을 효과적으로 찾기 위해서는 자막 영역에서 프로젝션의 peak가 강하게 나타나는 것이 바람직하다. 따라서 본 논문에서는 획 경계선의 내부를 채운 다음 프로젝션을 수행함으로써 프로젝션이 효과적으로 수행될 수 있도록 하였다. 이를 위하여 모폴로지 연산을 다시 한번 수행하여 인접한 획 경계선들을 통합한다. 획 경계 영역에 대해 모폴로지 폐쇄 연산을 수행하면 서로 인접한 획 경계는 자막 영역으로 병합되며, 획 경계 내부 영역도 채워지게 된다. 그러나 획 영역과 유사한 구조를 갖는 배경이 있는 경우에는 획 영역은 자막으로 병합되지 않고 유지된다. 그런 다음 수평, 수직 방향의 프로젝션을 반복적으로 수행하면, 잡음이나 배경에 의한 획 영역을 효과적으로 제거할 수 있다. 남아있는 영역들 중 기하학적으로 자막 영역이 되기에는 너무 크거나 작은 영역은 제거하여 최종적으로 자막 영역을 생성한다. 이 영역들은 이전 프레임의 자막 영역들과의 정합을 통해 추적된다.

III. 자막 추적

3.1 개요

각 프레임에서 추출된 자막 영역은 바운딩박스로 표현되어 인접한 프레임의 바운딩 박스와의 정합을 통해 추적된다. 자막 추적 단계에서는 이전 프레임에 존재하는 바운딩박스들의 목록인 바운딩박스 시퀀스 BS(t-1)와 현재 프레임에 있는 바운딩박스 B(t)들을 입력으로 받아, 바운딩박스 매칭을 통해 자막을 추적한다. BS(t-1)은 t-1프레임에서 관리, 추적하고 있는 바운딩박스의 모든 정보를 가지고 있는데, 바운딩박스가 시작된 프레임부터 현재까지 바운딩박스의 위치, 크기, 밝기 정보 등을 갖는다. 본 연구에서 매칭을 위해 사용한 특징(feature)은 바운딩박스의 위치, 크기, 그리고 밝기 정보 등이다.

현재 프레임에 대한 바운딩박스 추출과 기하학적 검증이 이루어진 후, 바운딩박스 시퀀스 BS(t-1)에 존재하는 바운딩박스들과 현재 프레임에서 얻은 바운딩박스 B(t)들과의 대응관계를 검사한다. 인접한 프레임에서 동일한 텍스트를 나타내는 바운딩박스는 그 크기와 밝기 구조가 서로 같아야 한다. 그러나 잡음과 복잡한 배경, 그리고 테두리 효과 등의 영향으로 서로 인접한 프레임에 존재하는 동일 텍스트 영역에 대한 바운딩박스는 그 크기와 좌표가 정확하게 일치하지 않는 경우가 대부분이고 이럴 경우 바운딩박스에 대한 정합이 어렵다. 이러한 문제를 해결하기 위하여 본 연구에서는 텍스트 추적을 위하여, 바운딩박스 탐색, 바운딩박스 정합, 바운딩박스 시퀀스 생성으로 이루어진 3단계 알고리즘을 제안한다.

3.2 바운딩박스 탐색

바운딩박스 탐색단계에서는 바운딩박스 시퀀스 BS(t-1, j)에 있는 바운딩박스와 B(t, i)에 있는 바운딩박스 사이의 상관성을 검사하여 BS(t-1, j)에 존재하는 바운딩박스가 대응될 가능성이 있는지를 검사한다. 여기에서 BS(t-1, j)는 t-1번째 프레임에 존재하는 j 번째 바운딩박스 시퀀스를, B(t, i)는 t번째 프레임에 존재하는 i번째 바운딩박스를 각각 의미한다.

서로 인접한 프레임에 존재하는 동일 텍스트 영역에 대한 바운딩박스의 크기는 잡음과 복잡한 배경, 그리고 테두리 효과 등의 영향으로 그 크기와 좌표가 정확하게 일치하지 않는 경우가 많다. 본 논문에서는 바운딩박스 시퀀스 BS(t-1, j)에 있는 바운딩 박스와 위의 자막 탐색 영역 내에 존재하는 t번째 프레임의 바운딩 박스가 충분히 겹쳐지면 이 바운딩박스는 서로 대응될 가능성이 있다고 판단하고 바운딩박스 정합단계를 수행한다.

그림 2는 이전 프레임까지의 바운딩박스 시퀀스 BS(t-1, j)와 그에 대응하는 현재 프레임 B(t, i)의 바운딩박스의 예를 보이고 있는데, 그 크기와 위치가 정확하게 일치하지 않는 모습을 보이고 있다. 본 논문에서는 두 바운딩박스의 겹쳐지는 영역이 일정 비율 이상일 경우 서로 대응될 가능성이 있다고 판단한다. 겹침 비율 은 (식 1)-(식 4)에 의해 계산된다.

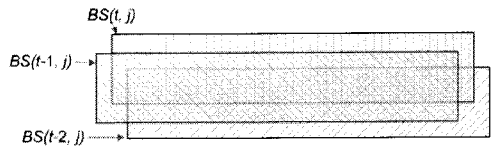


그림 2 완전히 겹쳐지지 않는 바운딩박스의 예
Fig. 2. An example of incompletely overlapped bounding box

$$R_{t-1} = (x_2^{t-1} - x_1^{t-1})(y_2^{t-1} - y_1^{t-1}) \quad \text{(식 1)}$$

$$R_t = (x_2^t - x_1^t)(y_2^t - y_1^t) \quad \text{(식 2)}$$

$$O = (MIN(x_2^{t-1}, x_2^t) - MAX(x_1^{t-1}, x_1^t)) * (MIN(y_2^{t-1}, y_2^t) - MAX(y_1^{t-1}, y_1^t)) \quad \text{(식 3)}$$

$$O_r = w_1 \frac{O}{R_t} + w_2 \frac{O}{R_{t-1}} \quad \cdots\cdots \text{(식 4)}$$

여기에서 (x_1^t, y_1^t), (x_2^t, y_2^t)은 t번째 프레임에 존재하는 바운딩박스의 좌상(left-top) 좌표와 우하(right-bottom) 좌표를 의미하고, R_t는 t번째 프레임에 존재하는 바운딩박스의 면적, O는 겹쳐지는 영역의 면적, 그리고 w_1, w_2은 가중치로서 $w_1 + w_2 = 1$이다. 본 논문에서는 이전 프레임에서 전달된 BS(t-1, j)의 모든 바운딩박스시퀀스에 대하여 현재 프레임의 모든 바운딩박스 B(t, i) 와 겹쳐짐 비율 R_r을 계산하여 R_r이 임계치 이상인 바운딩박스 B(t, i)는 바운딩박스 정합 단계로 넘기고 임계치 이하인 바운딩박스는 새로운 바운딩박스 시퀀스로 하여 BS(t)에 등록한다.

3.3 바운딩박스 정합

탐색 영역 내에 존재하는 바운딩박스 중 겹침 비율 R_r이 충분히 큰 바운딩박스에 대해 바운딩박스 정합을 수행하는데, 바운딩박스 정합 척도는 계산량이 적은 MAD(Mean of Absolute Difference)를 사용하였다. (식 5)는 MAD를 계산하기 위한 식을 보이고 있다.

Mean of Absolute Difference =

$$\frac{1}{n}\sum_{x}\sum_{y}|I_t(x,y) - I_{t-1}(x,y)| \quad \cdots\cdots \text{(식 5)}$$

여기에서 $I_t(x,y)$는 t번째 프레임의 (x, y) 좌표의 밝기 값을, n은 바운딩박스 내 픽셀의 갯수를 각각 의미한다. 그러나 그림 2에서 보인 바와 같이 두 바운딩박스의 크기가 일치하지 않는 경우가 많다. 따라서 본 논문에서는 두 바운딩박스의 크기가 정확히 일치하지 않은 경우, 두 바운딩박스가 모두 겹쳐지는 영역에 대해서만 MAD를 계산하고, MAD가 충분히 작은 경우 두 바운딩박스가 서로 대응되는 것으로 하였다.

3.4 바운딩박스 시퀀스 생성

현재 프레임에서 찾아진 바운딩박스들은 이전 프레임까지의 바운딩박스들을 관리하는 바운딩박스 시퀀스 BS(t-1, j)에 있는 바운딩박스들과 바운딩박스 탐색과 바운딩박스 정합을 수행한다. 정합 결과 다음 3가지 중 한 개의 결과를 얻을 수 있다.

i) BS(t-1, j)의 바운딩박스와 정합되는 바운딩박스를 B(t)에서 찾은 경우

ii) BS(t-1, j)의 바운딩박스와 정합되는 바운딩박스를 B(t)에서 찾지 못한 경우

iii) B(t, i)에는 존재하지만 BS(t-1)에는 대응되는 바운딩박스가 존재하지 않는 경우

i) 인 경우는 이전 프레임에 존재하던 텍스트가 현재 프레임에서도 발견된 경우이기 때문에 BS(t-1, j)에 B(t, i)에서 찾은 바운딩박스를 추가하여 현재 프레임의 바운딩박스 시퀀스를 BS(t, j)를 생성한다. ii)인 경우는 이전 프레임까지는 존재했던 텍스트가 현재 프레임에서 없어진 경우와 현재 프레임에 존재하지만 어떤 이유로 인해 대응관계를 찾는데 실패한 경우이다. 전자일 경우에는 BS(t-1)에서 해당 바운딩박스의 추적을 종료해야 하며, 후자일 경우에는 BS(t)에 BS(t-1)의 바운딩박스를 추가하지는 않되, B(t+1)의 바운딩박스와 대응관계를 탐색하기 위해 해당 바운딩박스를 다음 프레임으로 전달시켜야 한다. 본 논문에서는 현재 프레임 t에서 대응관계를 찾지 못한 바운딩박스 시퀀스는 이전 프레임 t-1에 존재하는 바운딩박스를 현재 프레임으로 복사하여 t 프레임으로 전달한 후 t+1 프레임과의 대응관계를 검사 할 수 있도록 한다. 이렇게 t 프레임으로 복사된 바운딩박스는 t+1에서도 대응 관계를 찾지 못할 경우에는 사라진 바운딩박스로 간주하여 바운딩박스에서 제거하며, t+1에서 대응관계를 찾을 경우에는 어떤 이유로 인해 일시적으로 대응관계를 찾지 못한 것으로 간주하여 이후 프레임에서 정상적으로 바운딩박스에 대한 추적을 계속한다. 마지막으로 iii) 인 경우는 새롭게 나타나는 텍스트를 의미하며 BS(t)에 새로운 바운딩박스 시퀀스로 등록한다. 현재 프레임에 존재하는 바운딩박스에 대한 정합이 모두 끝나면 BS(t)가 생성되어 추적 중인 바운딩박스의 시퀀스를 t+1 프레임에 전달하고, 한편으로는 텍스트가 끝나는 것으로 판정된 바운딩박스에 대해서는 자막 향상 과정을 계속 수행한다.

3.5 시간 중복성을 이용한 자막 향상

현재 프레임에 대한 바운딩박스 정합이 끝나면 새로운 바운딩박스 시퀀스 BS(t)가 생성되는데, BS(t)에는 BS(t-1)의 바운딩박스와 정합되는 바운딩박스를 B(t)에서 찾은 경우, BS(t-1)의 바운딩박스와 정합되는 바운딩박스를 B(t)에서 찾지 못한 경우 그리고, B(t)에는 존재하지만 BS(t-1)에는 대응되는 바운딩박스가 존재하지 않는 경우를 모두 포함한다. 이 중 BS(t-1)의 바운딩박스와 정합되는 바운딩박스를

B(t)에서 찾지 못한 경우는 BS(t-1)의 바운딩박스를 BS(t)로 복사한 것이다. 이 경우는 이전 프레임까지는 존재했던 텍스트가 현재 프레임에서 사라져 더 이상 존재하지 않는 경우와, 현재 프레임에 존재하지만 어떤 이유로 인해 대응관계를 찾는데 실패한 경우를 포함하는데, 바운딩박스가 더 이상 존재하지 않는 경우는 BS(t)의 바운딩박스 시퀀스와 BS(t-1)의 시퀀스는 각각 BS(t-1)과 BS(t-2)에서 복사한 것이다. 따라서 본 논문에서는 바운딩박스 시퀀스 중 BS(t)와 BS(t-1)이 모두 복사된 시퀀스인 경우 사라진 자막으로 판단하고 자막 향상과정을 수행한다.

한편 디지털 비디오의 자막은 시청자가 자막을 읽을 수 있도록 일정 프레임 이상 존재 한다. 따라서 사라진 바운딩박스 중 일정 시간 이상 추적되지 않은 경우에는 정상적인 바운딩박스라고 할 수 없다. Palma 등의 관찰에 의하면 정상적인 자막 정보는 비디오에서 적어도 0.25초 이상 존재하며[17], 이에 따라 본 논문에서는 0.25초 이상 추적된 바운딩박스인 경우에는 정상적인 텍스트로 간주하여 자막 향상 및 자막 추출 단계를 수행한다. 추적이 끝난 바운딩박스 시퀀스는 적어도 0.25초 동안 5프레임 이상을 갖는다. 따라서 시간적 중복성을 이용한 자막 향상이 가능하다. 그러나 바운딩박스 시퀀스의 바운딩박스들은 여러 가지 잡음의 영향과 더불어 자막 바운딩박스 탐색과정에 자막에 대한 여러 가지 특수효과로 인해 그 크기와 위치가 정확하게 일치하지 않는 경우가 대부분이다. 본 논문에서는 서로 일치하지 않는 위치의 바운딩박스 시퀀스의 최종 위치 결정은 해당 바운딩박스 시퀀스의 크기와 위치에 대한 누적된 정보를 기반으로 한다. 바운딩박스의 최종 위치 (x1, y1), (x2, y2)는 다음의 식에 의해 계산된다.

$$(x1, y1) = \left(\frac{1}{n}\sum_{i=1}^{n} x1_i, \frac{1}{n}\sum_{i=1}^{n} y1_i \right) \cdots\cdots\cdots\cdots \text{(식 6)}$$

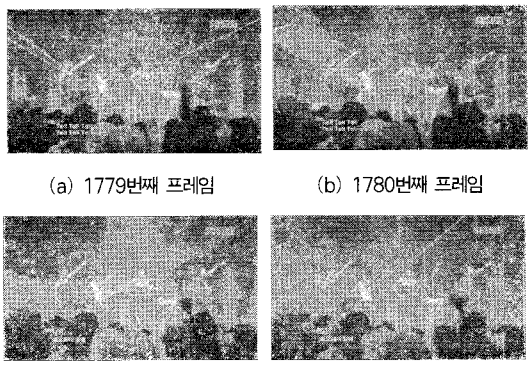
$$(x2, y2) = \left(\frac{1}{n}\sum_{i=1}^{n} x2_i, \frac{1}{n}\sum_{i=1}^{n} y2_i \right) \cdots\cdots\cdots\cdots \text{(식 7)}$$

여기에서 (x1, y1)는 최종 바운딩박스의 좌상 좌표를, (x2, y2)는 우하 좌표를 각각 의미한다.

(식 6)-(식 7)에 의해 계산된 최종 바운딩박스 위치에 대해서 바운딩박스 시퀀스에 대해 시간적 중복성(redundancy)를 이용한 자막 영역 향상을 수행한다. 시간축을 기준으로 텍스트 영역의 밝기는 거의 변하지 않는 반면 배경 영역은 변하기 때문에 이 과정을 수행한 영역은 자막 영역의 화질은 더욱 향상되는 반면 배경 영역은 흐릿하게(blurring)되기 때문에 이진화를 통한 자막 추출에 적합하게 된다.

IV. 실험결과

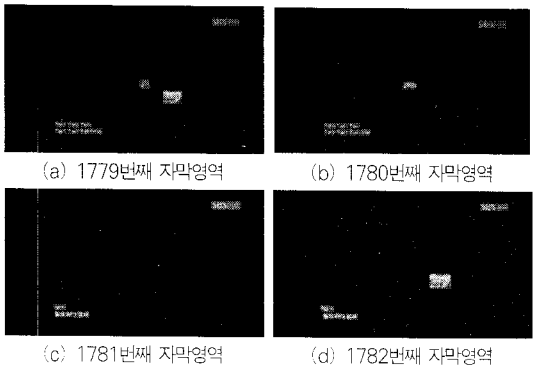
본 논문에서 제안하는 자막 추출 기법의 타당성을 입증하기 위하여 다양한 비디오에 대한 실험을 수행하였다. 실험에 사용한 영상은 720*480 또는 960*540 해상도, 256단계의 농담(gray scale)값을 갖는 뉴스, 스포츠, 쇼 프로그램 등이다. 그림 3은 실험에 사용한 영상열의 일부를 보이고 있는데, 그림 3(b)와 그림 3(c)에서 자막의 내용이 바뀌고 있는 모습을 보이는 것을 볼 수 있다. 그림 4는 그림 3의 입력 시퀀스에 대해 참고문헌 [7]에서 제안한 방법으로 추출한 자막 영역을 보이고 있는데, 그림에서 보이는 바와 같이 잘못된 2개의 자막 영역을 포함하여 총 5개의 자막 영역이 검출되고 있는 것을 볼 수 있다. 그림 4에서 찾아진 자막 영역들은 인접한 프레임의 자막 영역들과 정합을 통하여 추적되는데, 그림 4(b)와 그림 4(c)에서 보이는 자막 영역은 서로 다른 텍스트 정보를 나타내고 있기 때문에 그림 4(b)에서 추적이 중단되고 다시 그림 4(c)에서 새로운 자막 영역에 대한 추적을 시작한다.



(a) 1779번째 프레임 (b) 1780번째 프레임

(c) 1781번째 프레임 (d) 1782번째 프레임

그림 3. 입력 시퀀스의 예
Fig. 3. Examples of input sequence



(a) 1779번째 자막영역 (b) 1780번째 자막영역

(c) 1781번째 자막영역 (d) 1782번째 자막영역

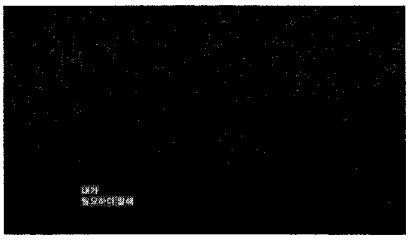
그림 4. 자막 영역 추출 결과
Fig. 4. The results of text region extraction

한편 1779번째 프레임에 존재하지만 1780번째 프레임에는 존재하지 않는 바운딩박스는 바운딩박스 시퀀스 길이가 5 프레임 미만이기 때문에 잘못된 바운딩박스로 판단하고 버린다. 추적이 중단된 자막 영역 중 의미 있는 자막 영역은 시간적 평균화를 통해 그 화질을 향상시키는데, 그림 5는 그 결과를 보이고 있다. 그림 5(a)는 1782번째 프레임에서 추적에 실패한 바운딩박스 시퀀스 중 그 길이가 5프레임 이상인 바운딩박스의 화질을 향상시켜 출력한 것이다. 그림 5(b)는 10 프레임 이상 계속 추적되고 있는 바운딩박스를 출력한 것으로써 1784번째 프레임에서 출력된 것이고 그림 5(c)는 1791번째 프레임에서 출력된 것이다.

제안된 방법을 방송 프로그램에서 추출한 981프레임에 존재하는 5466개의 글자에 대하여 실험한 결과 잘못된 텍스트 검출(false positive)은 없었으며 텍스트를 추출하는데 실패한 경우(false negative)는 508회가 관측되어, 본 논문에서 제안된 알고리즘의 정확도(precision)는 100%, 검출률(recall)은 90.7% 이었다.

(a) 1781번째 프레임에서 추출된 자막

(b) 1784번째 프레임에서 추출된 자막



(c) 1791번째 프레임에서 추출된 자막

그림 5. 최종 결과
Fig. 5. Final results

V. 결론

본 논문에서는 시공간 정보를 이용하여 텍스트 정보를 효과적이고 강건하게 추출하기 위한 알고리즘을 제안하였다. 제안된 알고리즘은 각 프레임에 대해 자막이 될 가능성이 높은 픽셀을 의미하는 씨앗점을 추출하고 모폴로지 연산과 프로젝션 기법을 이용하여 자막 영역을 찾는다. 그런 다음 시간적으로 인접한 프레임간의 자막 영역을 MAD를 이용하여 추적한다. 추적이 끝난 자막 영역은 시간적 중복성을 이용하여 화질을 향상시켜 최종 자막을 추출한다. 제안된 방법은 획의 특성에 기반하여 자막 영역를 추출하고, 일부 프레임에서 자막 추적에 실패하더라도 강건하게 자막 영역 추적을 계속할 수 있으며 이는 실험을 통해 확인되었다. 향후, 자막 영역에 대한 이진화 방법에 대한 연구와 이진화된 자막 영역 중 어느 영역이 자막이고 어느 영역이 배경인지 판단하기 위한 색상 극화(color polarity) 기법에 대한 연구가 필요하다.

참고문헌

[1] 신성윤, 표성배, 이양원, "대용량 비디오 데이터베이스 구축을 위한 비디오 개요 추출," 한국컴퓨터정보학회논문지, 제14권 1호. 255-265쪽, 2006년.

[2] K. Jung, K. Kim, A. Jain, "Text Information extraction in images and video: a survey," Pattern Recognition, vol. 37, pp. 977-997, May 2004.

[3] M. R. Lyu, J. Song, and M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction," IEEE Trans. on CSVT., vol. 15n no 2, Feb. 2005.

[4] R. Lienhart and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," IEEE Trans. on Circuits and Systems for Video technology, vol. 12, no. 4, pp. 256-268, Apr. 2002.

[5] Y. Hasan and L. Karam, "Morphological Text Extraction from Images," IEEE Trans. on Image Processing, vol. 9, no. 11, Nov. 2000.

[6] C. W. Lee, K. Jung, H. J. Kim, "Automatic text detection and removal in video sequences," Pattern Recognition Letters, vol. 24, pp. 2607-2623, Nov. 2003.

[7] 정종면, 차지훈, 김규헌, "디지털 비디오를 위한 획기반 자막 추출 알고리즘," 퍼지 및 지능시스템학회 논문지, vol. 17, no. 3, pp. 297-303, 2007년 6월

[8] H. Li, D. Doerman, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. on Image Processing, vol. 9, no. 1, pp. 147-156, Jan. 2000.

[9] O. Shiku, Y. Xiao, H. Yan, "Extraction of character patterns in different styles and orientations from natural scene images," Proc. of 2004 Int. Symp. on Intelligent Multimedia, Video and Speech Processing, pp. 719-722, Oct. 2004.

[10] A. Jian and S. Bhattacharjee, "Text segmentation using gabor filters for automatic document processing," Machine Vis. Applicat., vol. 5, pp. 169-184, 1992.

[11] V. Wu, R. Manmatha, and E. Riseman, "Textfinder: An automatic system to detect and recognize text in images," IEEE Trans. on Pattern Analysis and Machine Intelligent, vol. 21, no. 11, pp. 1224-1229, Nov. 1999.

[12] A. Jain and B. Yu, "*Automatic text location in images and video frames*," Pattern Recognition, vol. 31, no. 12, pp. 2055-2076, 1998.

[13] M. Cai, J. Song, and M. Lyu, "*A new approach for video text detection*," Proc. of Int. Conf. on Image Process, pp. 117-120, Sep. 2002.

[14] A. Wernicke and R. Lienhart, "*On the segmentation of text in videos*," Proc. of IEEE Int. Conf. on Multimedia Expo, vol. 3, pp. 1511-1514, Jul. 2000.

[15] S. Antani, D. Crandall, and R. Kasturi, "*Robust extraction of text in video*," 15th Int. Conf. on Pattern Recognition, vol. 1, pp. 831-834, Sep. 2001.

[16] W. Huang, P. Shivakumara and C. L. Tan, "*Detecting Moving Text in Video Using Temporal Information*," Proceedings of 19th ICPR, Dec. 2008.

[17] D. Palma, J. Ascenso, F. Pereira, "*Automatic Text Extraction in Digital Video Based on Motion Analysis*," LNCS 3211 Image Analysis and Recognition, pp. 588-596, 2004.

저 자 소 개

정 종 면

1992 : 한양대학교 공학사
1994 : 한양대학교 공학석사
2001 : 한양대학교 공학박사
2001 - 2004
한국전자통신연구원 선임연구원
2008 - 2009
The Ohio State University
Visiting Scholar
2004 - 현 재
국립목포해양대학교
해양전자통신공학부 교수
관심분야 : 영상처리, 머신 비전,
디지털방송, MPEG-2, 4, 7, 21 응용